Niveau(x) d'enseignement: 0 Numéro EDUQ : 0478890
Centre d'intérêt: 570

Allaire, Denis.

La fidélité et la validité des questionnaires d'évaluation des professeurs : essai de reformulation / par Denis Allaire. -- [Montréal : PERPE, 197-].
5 f. ; 28 cm.

Document: 0573-02.

Descripteur(s): *Evaluation de l'enseignant *Evaluation par l'étudiant *Satisfaction de l'étudiant *Attitude de l'enseignant *Relations maître-élèves *Analyse statistique *Validité *Questionnaire *Test *Evaluation de tests

Identificateur(s) de nom: *PERPE (Test)

Un professeur qui reçoit ses résultats à un questionnaire d'évaluation des professeurs (QEP) s'interroge habituellement sur le degré de confiance qu'il doit leur accorder. En fait, de multiples facteurs sont susceptibles de contribuer à ces résultats et il faut connaître leur importance respective pour interpréter de façon nuancée les résultats de tels questionnaires. Ce document étudie l'ensemble de ces facteurs: 1. les facteurs variables [inconstance des réponses des étudiants(e), différences entre étudiants (j), occasion de l'évaluation (o), particularités attachées au cours (c) et expérience du professeur (t)]; 2. les facteurs fixes [item du questionnaire (i), formulation des questions (f), approche qu'adopte l'évaluation (a) et méthode d'observation (m)]. Cette étude de la généralité des scores réunit dans une même approche les études traditionnelles de validité et de fidélité.

Type de document: r

EDUQ 8470090

-73

La fidélité et la validité des Questionnaires d'Evaluation des Professeurs

Essai de reformulation

par

Denis Allaire

Document 0573-02

Un professeur qui reçoit ses résultats à un questionnaire d'évaluation des professeurs (QEP) s'interroge habituellement sur le degré de confiance qu'il doit leur accorder. Est-ce que ses résultats auraient été les mêmes si ses étudiants avaient recommencé l'évaluation le jour même? Si oui, est-ce qu'un autre groupe d'étudiants aurait fourni la même évaluation? Est-ce qu'elle aurait été la même la semaine avant ou la semaine après? A quel point l'évaluation faite dans un autre cours qu'il enseigne risquerait d'être tout autre? Est-ce que ses présents résultats risquent de ressembler à ceux qu'il obtiendrait dans quelques années? Est-ce que les résultats issus d'un questionnaire différent quant aux choix et à la formulation des questions seraient équivalents? Autrement dit, ses résultats le caractérisent-il vraiment? définitivement? Ou, au contraire, ne sont-ils guère plus valables qu'un horoscope?

L'inconstance des réponses des étudiants (que nous symboliserons "e"), les différences entre étudiants ("j" pour juges), l'occasion de l'évaluation à l'intérieur du cours ("o"), les particularités attachées au cours que livre le professeur ("c") ainsi que l'expérience particulière qui est la sienne à un moment donné de sa carrière ("t" pour temps) sont autant de facteurs sueceptibles de contribuer aux scores qu'il obtient.

Ce n'est qu'à la condition de connaître leur importance respective qu'il est possible d'interpréter de façon nuancée les résultats issus de pareils questionnaires. Par exemple, si le facteur "étudiants" (j) est important, c'est dire que la mesure obtenue risquerait d'être tout autre si le professeur avait donné son cours à un autre groupe. Dans un tel cas, l'interprétation devra en tenir compte. La mesure obtenue étant largement fonction d'un groupe d'étudiants, la généralisation qu'on peut en faire ne devrait pas porter sur les autres groupes. Une mesure permettant une généralisation aussi restreinte pourra quand même être utile si le but du professeur est de connaître l'opinion de <u>ses</u> étudiants de façon

à s'adapter le plus possible à leurs exigences. Si, par contre, son intérêt est de savoir à quel point son enseignement pouvait répondre aux attentes de la moyenne des étudiants, cette impossibilité de généraliser au-delà de son groupe précis d'étudiants empêchera d'attacher foi aux mesures obtenues.

De même pour le facteur "occasions". Si l'opinion qu'ont les étudiants de leur professeur est sujette à de grandes variations d'une occasion à l'autre à l'intérieur du même cours, il deviendra impossible de généraliser la mesure issue d'une seule occasion à l'ensemble du cours. En supposant également qu'il existe de grandes différences entre les différents cours que peut donner un même professeur à une même session, on pourra difficilement se servir de la mesure issue d'un seul cours pour généraliser à ce que serait sa performance dans les autres cours. De même du facteur "temps", dont l'importance risque d'empêcher toute utilisation pronostique des résultats provenant d'une seule session dans la carrière du professeur. Cette impossibilité de généraliser sur le temps conservera l'utilité diagnostique des résultats, mais empêchera peut-être un employeur de s'en servir dans l'attribution d'un poste à long terme.

Nous symboliserons donc le score d'un professeur "p" par XpoctJ[1] pour rendre explicite tous les facteurs, autres que P, qui sont susceptibles d'influence. Ces facteurs étant confondus avec le facteur "professeurs", les différences observées d'un professeur à l'autre pourraient théoriquement être explicables par n'importe lequel. Ces facteurs sont dits "confondus" et non "fixes" parce que pour chaque nouveau professeur évalué correspond une nouvelle condition de chacun d'eux. A la différence entre deux résultats correspond, non pas seulement deux professeurs différents, mais également deux occasions différentes, deux groupes de juges différents, deux cours différents, deux moments différents à l'intérieur de leur carrière respective.

---

[1] Le J majuscule indique que plusieurs juges, et non pas un seul, ont contribué au score.

Il en est tout autrement du facteur que constituent les items (i) du questionnaire puisque tous les professeurs ont été soumis aux mêmes items. Ce facteur est dit "fixe" en ce sens qu'à chaque nouveau professeur ne correspond pas un nouveau choix d'items. Un facteur fixe contribue également aux différences observées entre professeurs, mais son influence doit être distinguée de celle des facteurs confondus. Ainsi une différence entre les scores de deux professeurs ne peut s'expliquer par le fait que les items de l'un étaient plus faciles, c'est-à-dire de ceux qui provoquent en général le plus de satisfaction chez les étudiants, alors qu'elle peut très bien s'expliquer par le fait que les juges de l'un étaient moins sévères que la moyenne.

Cependant, le choix des items est quand même susceptible d'influence puisque le classement des professeurs pourrait être différent si le questionnaire comportait d'autres items. C'est le cas quand les items sont hétérogènes, quand il existe une interaction items-professeurs (pi). S'il s'avère que cette interaction est importante, la généralisation est alors limitée aux items du questionnaire.

Un autre facteur fixe intéressant à considérer est la formulation des questions (f). Tous les professeurs sont soumis à la même formulation qu'elle soit directe, indirecte ou autrement. Il est possible qu'une autre formulation modifie leur classement. Si tel est le cas, la généralisation sera limitée à une formulation donnée.

Un autre facteur serait l'approche (a) qu'adopte l'évaluation des professeurs. Utilise-t-on des étudiants, des collègues, le professeur lui-même, des administrateurs ou bien des spécialistes en pédagogie pour juger des professeurs? La généralisation qu'on fera de la mesure obtenue ne pourra être indépendante de ce facteur qu'à la condition que le classement demeure le même pour chacune des catégories de juges.

Un dernier facteur susceptible d'influence serait la méthode d'observation (m) sur laquelle se base la mesure. Les items peuvent être de type "comportemental" (juger de la fréquence d'un comportement obser-

vable) ou "introspectif" (juger un état subjectif, tel le sentiment d'avoir beaucoup appris). A quel point peut-on généraliser sur ce facteur? Autre aspect: l'observation peut être directe (présence en classe) ou différée (utilisation du vidéo, par exemple). A quel point, le classement des professeurs serait indépendant du choix de l'une ou l'autre méthode? Autrement dit, à quel point peut-on généraliser sur ce facteur?

Pour tenir compte explicitement de l'influence possible de ces facteurs fixes sur le score observé, nous le symboliserons désormais par XpoctJ|Ifam. Les lettres à droite du trait vertical représentent donc les facteurs fixes, tandis que les lettres à gauche représentent les facteurs confondus.

Toutes les questions que nous avons posées jusqu'ici à propos du score observé d'un professeur se résument en une question bien simple. Ce score est-il dépendant de certaines variables (o,c,t,j,e et i,f,a,m) qui pourraient en limiter la généralisation? L'ensemble des questions abordées ici nous apparaissent délimiter le champ de ce que devrait être l'étude psychométrique de ce genre de mesure (QEP). Ce qui a été proposé en ces pages est, somme toute, une étude de la généralité des scores XpoctJ|Ifam, étude qui réunit dans une même approche les études traditionnelles de validité et de fidélité. Si nous préférons le terme de généralité à ceux plus courants de fidélité et de validité, c'est qu'il souligne la continuité fondamentale entre ces deux terrains d'étude. Pour nous, la fidélité était l'étude de l'importance des facteurs confondus et la validité, l'étude des facteurs fixes pour une mesure donnée. En plus d'unifier ces deux domaines, une approche, comme celle proposée ici, dont le thème est l'étude de la généralité d'une mesure, a l'avantage de clarifier les liens entre diverses techniques d'analyse dans l'un et l'autre domaine. L'obscurité des liens, par exemple, entre diverses techniques de fidélité (retest immédiat, consistence interne, "split-half", retest lors d'une autre occasion du même cours, lors d'un autre cours, lors d'une autre session) disparaît dès que sont rendues explicites

les diverses variables sur lesquelles portent la généralisation. De même, cette approche permet de classifier les diverses études qui peuvent se réclamer de la validité puisqu'elle rend explicite les différents facteurs sur lesquels on tente de généraliser.

En guise de conclusion, le tableau 1 fait la synthèse des études de généralité énumérées jusqu'ici. Pour chacune, on retrouve sur quels facteurs porte la généralisation ("généralisation sur"), à quels facteurs elle se limite ("généralisation à") et quelle est la corrélation[1] permettant d'estimer le coefficient de généralité recherché.

---

[1] A noter que l'analyse de variance peut aussi bien servir à estimer le coefficient de généralité que le calcul d'une corrélation. De plus, certains plans d'analyse de variance permettent d'estimer plusieurs coefficients en une seule fois.

Tableau 1

Résumé des études de généralité
proposées pour un QEP

Généralisation à	Généralisation sur	Coefficient de généralité
I - Fidélité		
Mp|Ifam	o,c,t,J,e	corrélation entre 2 cours à des sessions différentes
Mpt|Ifam[a]	o,c,J,e	corr. entre 2 cours de la même session
Mptc|Ifam	o,J,e	corr. entre 2 demies de la classe en 2 occasions
MptcJ|Ifam	o,e	corr. entre deux occasions du même cours et avec les mêmes étudiants
Mptoc|Ifam[a]	J,e	corr. entre les demies de la classe lors de la même occasion
MptocJ|Ifam	e	corr. entre deux passations à la même occasion et avec les mêmes étudiants (retest immédiat)
II - Validité		
MpoctJ|fam[a]	I,e	corr. entre les deux demies des items (coefficient alpha)
MpoctJ|Iam[a]	f,e	corr. entre deux formulations de la question
Mpoct|Iam	J,f,e	corr. entre deux formulations chacune étant répondue par une demie de la classe
MpoctJ|Ifm	a,e	corr. entre deux approches à la rédaction des items
MpoctJ|Ifam	m,e	corr. entre deux catégories de juges
etc...		

[a] Ces études ont déjà été réalisées dans le cas du test PERPE.